Assignment 1

Ivan Aristy — iae225

October 2, 2024

1 Assignment

1.1 Designing a Static Visualization

Your task for this assignment is to design a static visualization that you believe effectively communicates an idea or message about the sunshine data, and provide a written report (no more than 1 page, single spaced, not including images) that details your design. Note that you should include at least one image of your visualization in your report (i.e., when all of the images are excluded from the report, the text does not exceed 1 page). We recommend that you start this assignment by identifying a question about the data that you would like your visualization to answer. Then, you should design a visualization to answer that question, and use that question as the title of your visualization.

You are required to use the given data set, but you are free to transform the data in any way that you wish. That is, you can manipulate the data using transformations such as a log transformation, computing percentages or averages, grouping elements into categories, or removing unnecessary variables. You are also allowed to incorporate additional data from external sources, but if you do so, you need to make a note of it in your submission.

Remember that different visualizations can emphasize different aspects of a data set, so your writeup should include details about which components of the data you intended to communicate. You should also provide comments on which aspects of the data set are obscured by your visualization design.

Your report should explain the rationale behind your design decisions. You should document the visual encodings you used and why they are effective for communicating your intended message. These decisions include things like the choice of visualization type, size, color, scale, and other visual elements (which we will learn more about during Week #3 and Week #4 lectures).

Note, the text for this document is under a standard word page when excluding the images and pretty formatting.

1.2   Intended Visualization & Question

What are the best cities for nice weather and lots of sunshine each month?

This is the question we aim to answer in this paper. In a more nuanced sense, the question can be interpreted as "What cities are between the 68 and 80°F range, while maximizing for sunshine hours?"

The reason for this question is that the 68-80°F range is considered the most comfortable for most people. The audience would be looking for the two best cities climate-wise for a vacation, to move to, or maybe a second house.

1.3   Data Cleaning & Gathering

We first need to consider if we have all the elements necessary to answer this question. In particular, we already have sunshine hours per month, and the particular cities we want to analyze. However, we do not have the temperature data for each city. Hence, we can retrieve this data from weather.gov
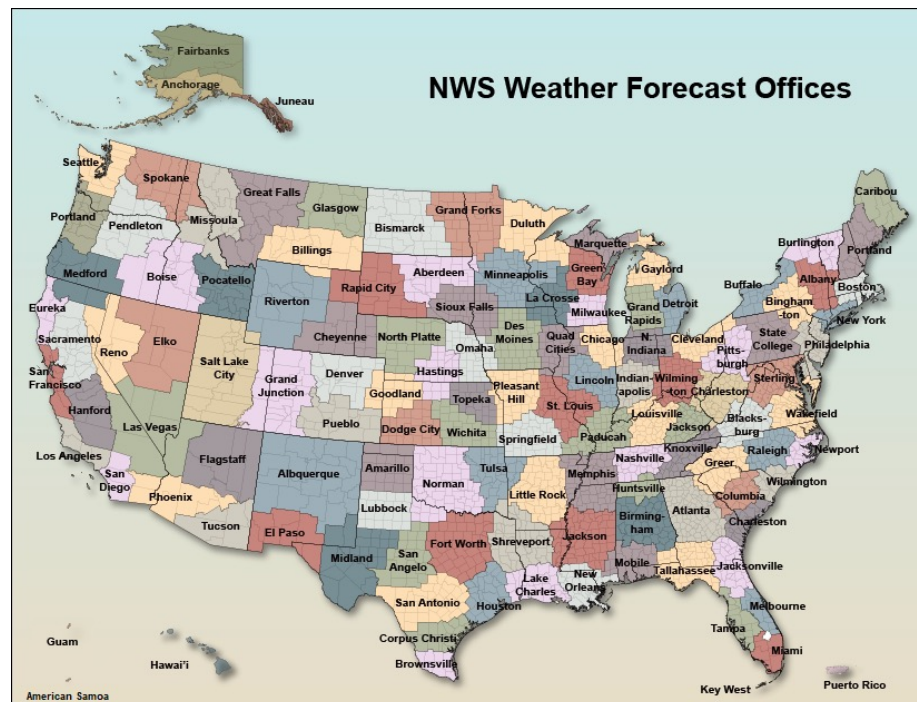
Figure 1: Map of the United States showing available data.

Figure 2: Data request for weather.gov

**Figure 3:** Obtained data for temperature.

Also, we do not need to consider latitude and longitude for our visualization, so we can remove or ignorethese columns.

1.4 Iterations

First I tried to filter the data to only show cities with a temperature between 68 and 80°F:
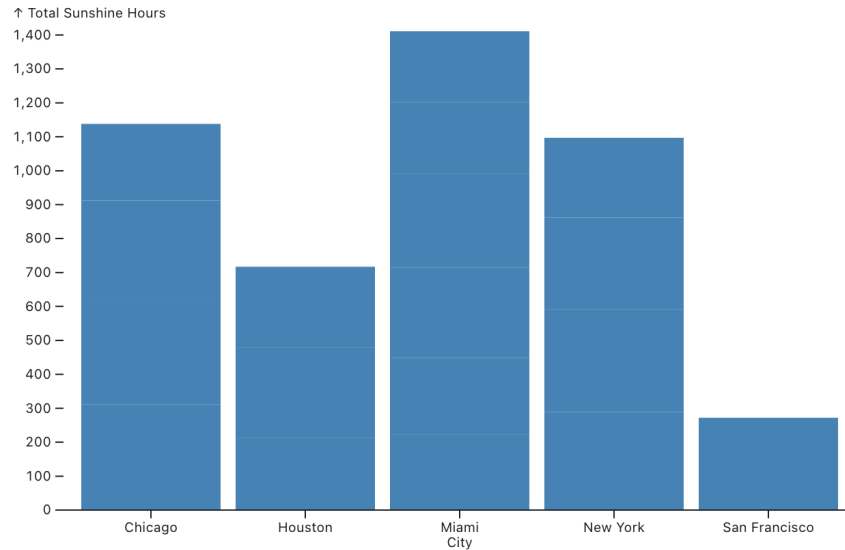
Ignore, that there's no visual encoding, this is just testing. However, I think this approach is not effective in communicating the data. We could instead try to encode the data using a scatter plot, with temperature on the x-axis, and sunshine hours on the y-axis.
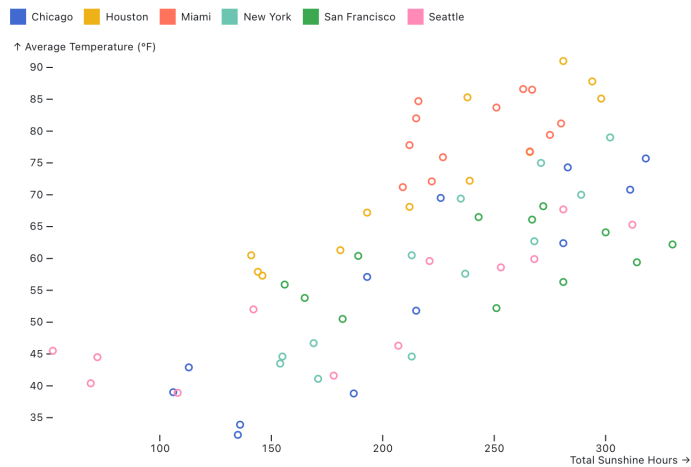
Adding a highlighted section for the 68-80°F range:
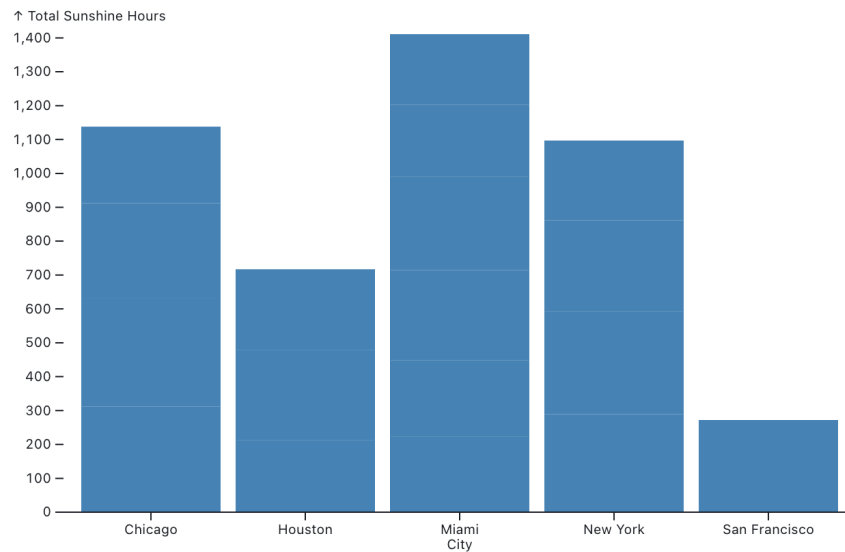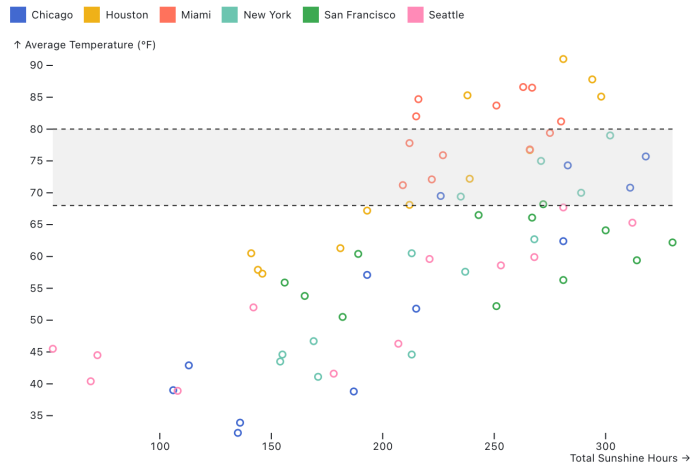
1.5 Final Visualization

1.5.1 Does this answer the question?

We highlight some key data:

- **Sunshine** lorem ipsum dolor sit amet

- **Temperature** lorem ipsum dolor sit amet

- **City** lorem ipsum dolor sit amet

1.5.2 Encoding (Marks/Channels, Redundancy, Color/Accessibilty)

1.5.3 Obscured Data

- **Preference:** The visualization is not dynamic, hence we cannot allow the user to select their desired temperature range.

- **Lat-Long:** Our visualization assumes that the audience is familiar with the cities in the dataset, and doesn't provide a map for reference.

- **Humidity:** We do not consider humidity in our visualization, which could be a factor in the comfort of the weather.